AMSR-E SIPS

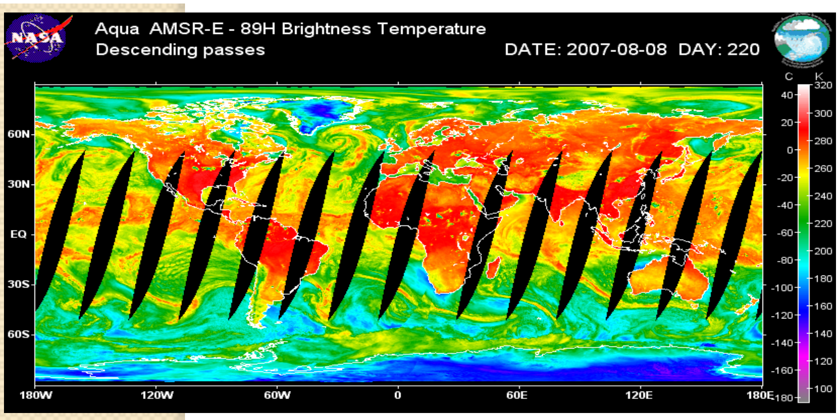
Helen Conover and Kathryn Regner

Information Technology & Systems Center

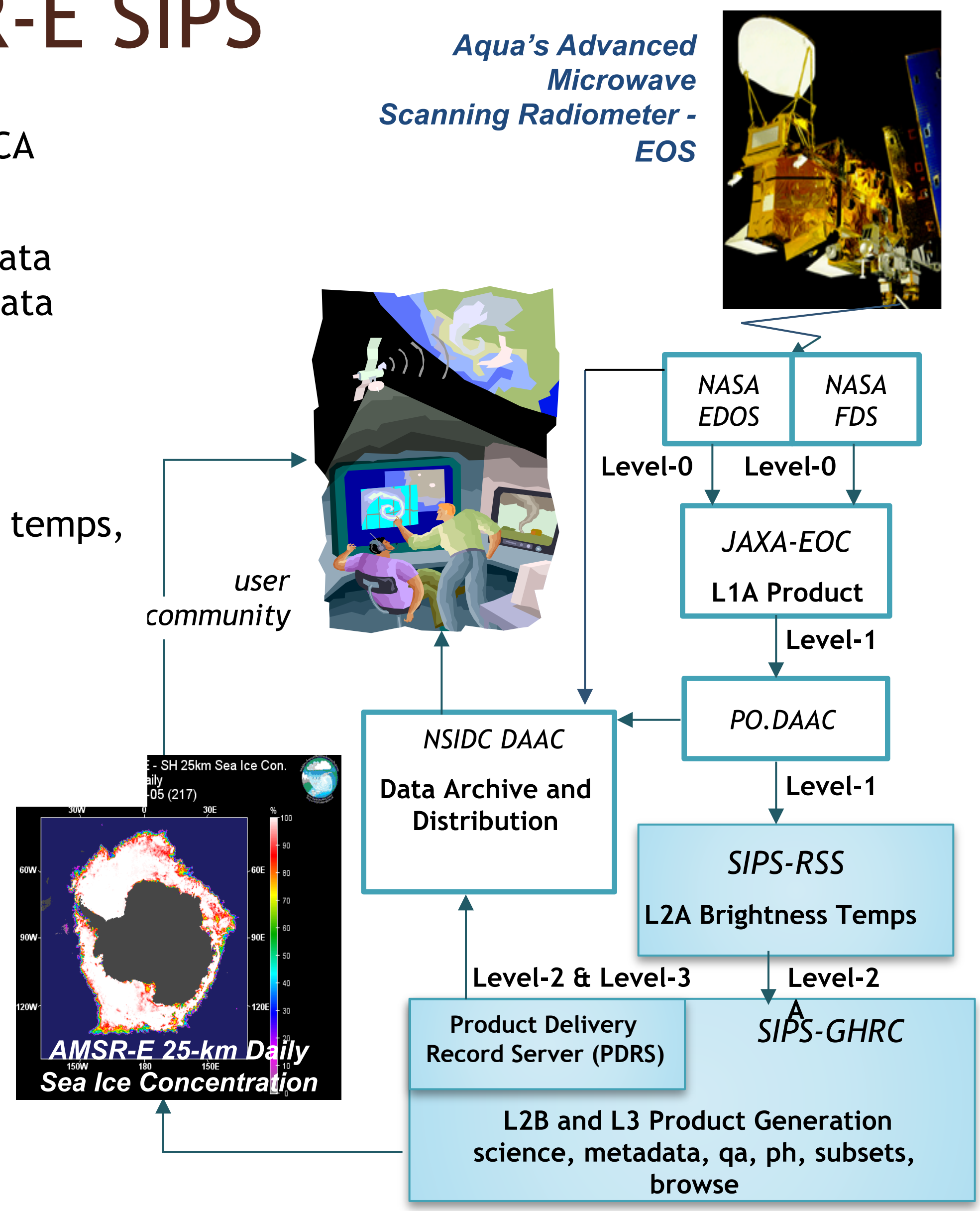
University of Alabama Huntsville

AMSR-E SIPS

- **Locations**: MSFC, Huntsville, Alabama
  Remote Sensing Systems, Santa Rosa, CA
- **Science Disciplines**: generating AMSR-E standard products for the global hydrologic cycle; science data subsets and imagery. Utilizing highly automated data transfers, processing and system monitoring
- **Science Team Size**: approximately 70, including instrument scientists and validation team
- **15 Products** covering rainfall, sea ice, sea surface temps, water vapor, cloud water, soil moisture, snow
- **Ingest rate** (approx): 2.5 GB/Day
- **Output Rate** (approx): 3.5 GB/Day
- Products **Archived** at: NSIDC DAAC



89H Brightness Temperatures Imagery Descending Passes for 2007-08-08



AMSR-E Data Flow

UAHuntsville
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

AMSR-E SIPS

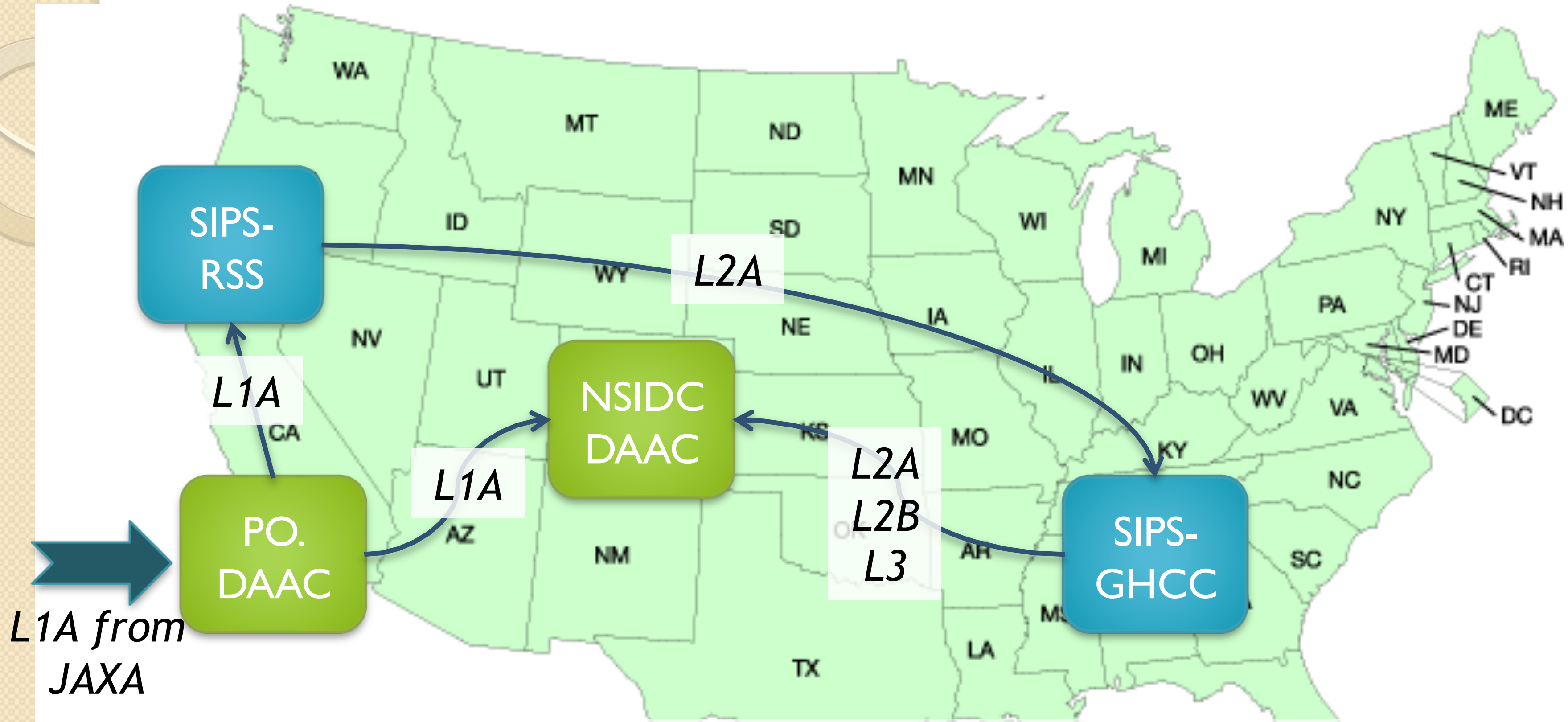
- Science Investigator-led Processing Systems (SIPS):
  - work in close cooperation with the instrument science teams and the Distributed Active Archive Centers (DAACs)
  - To generate high quality climate data products for use by Earth scientists and science application end-users

- The AMSR-E SIPS generates standard data products using algorithms developed by the U.S. science team and transfers those products to the National Snow and Ice Data Center (NSIDC) Distributed Active Archive Center (DAAC) where they are distributed to science researchers and the public.

AMSR-E Instrument and Data

- The Advanced Microwave Scanning Radiometer for EOS (*AMSR-E*) is a Japanese-designed passive microwave radiometer flying aboard NASA's *Aqua* satellite.
- Japanese Aerospace Exploration Agency (JAXA) generates Level 1 data and a suite of products
- AMSR-E SIPS generates NASA standard products from the Japanese L1A antenna temperatures
  - Level 2A Brightness temperatures
  - Level 2B derived products in native satellite projection (swath)
  - Level 3 gridded products (daily, pentad, weekly, monthly)
  - Custom subsets and imagery for the Science Team to support instrument validation and field campaigns

AMSR-E SIPS Distributed Architecture

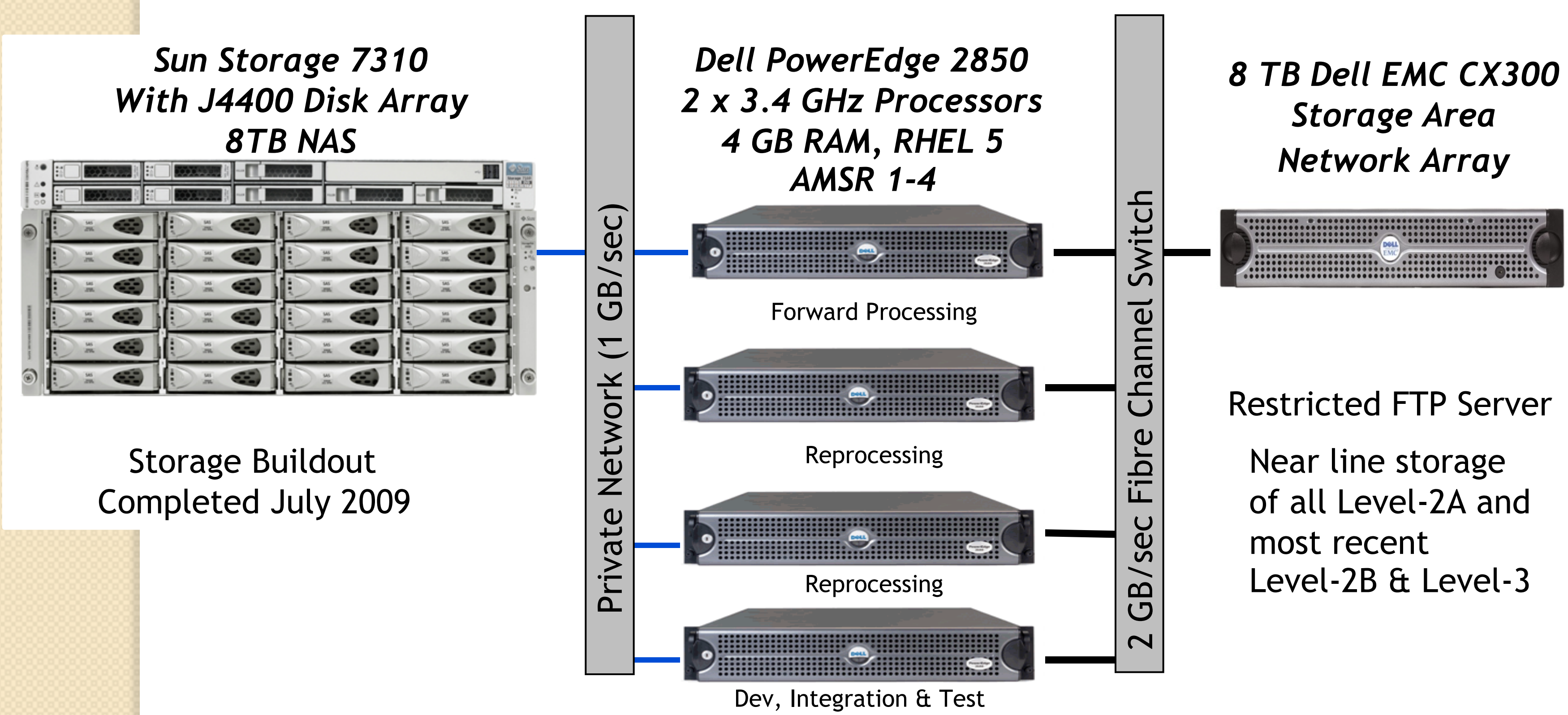


- The AMSR-E SIPS is a geographically distributed data processing system comprising two facilities located at
  - Remote Sensing Systems (RSS) in Santa Rosa, California
  - Global Hydrology Resource Center (GHRC) in Huntsville, Alabama

Interface Coordination

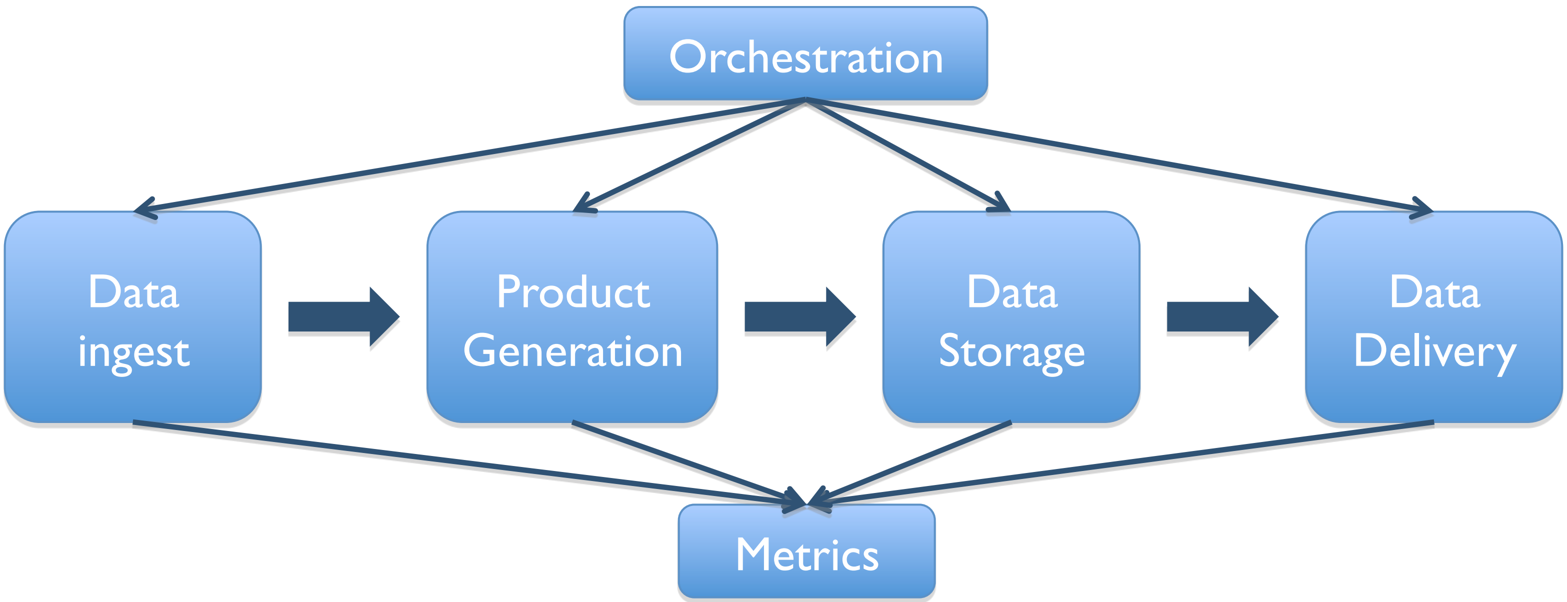
- AMSR-E SIPS working relationship with NSIDC DAAC is governed by:
  - Interface Control Documents
  - Operations Agreement
  - Data Management Plan
- Regular communication among ESDIS, Science Team, SIPS and DAAC
  - Bi-weekly telecons
  - Science team and ground systems meetings
  - Status and metrics reporting
- Aqua Mission Operations and Science Systems (MOSS) testing
  - Series of 7 tests over two years prior to launch
  - Exercising Aqua ground system with simulated data

SIPS-GHCC Hardware Operational Configuration



Multiple processing environments across four machines provide flexibility for processing and reprocessing.

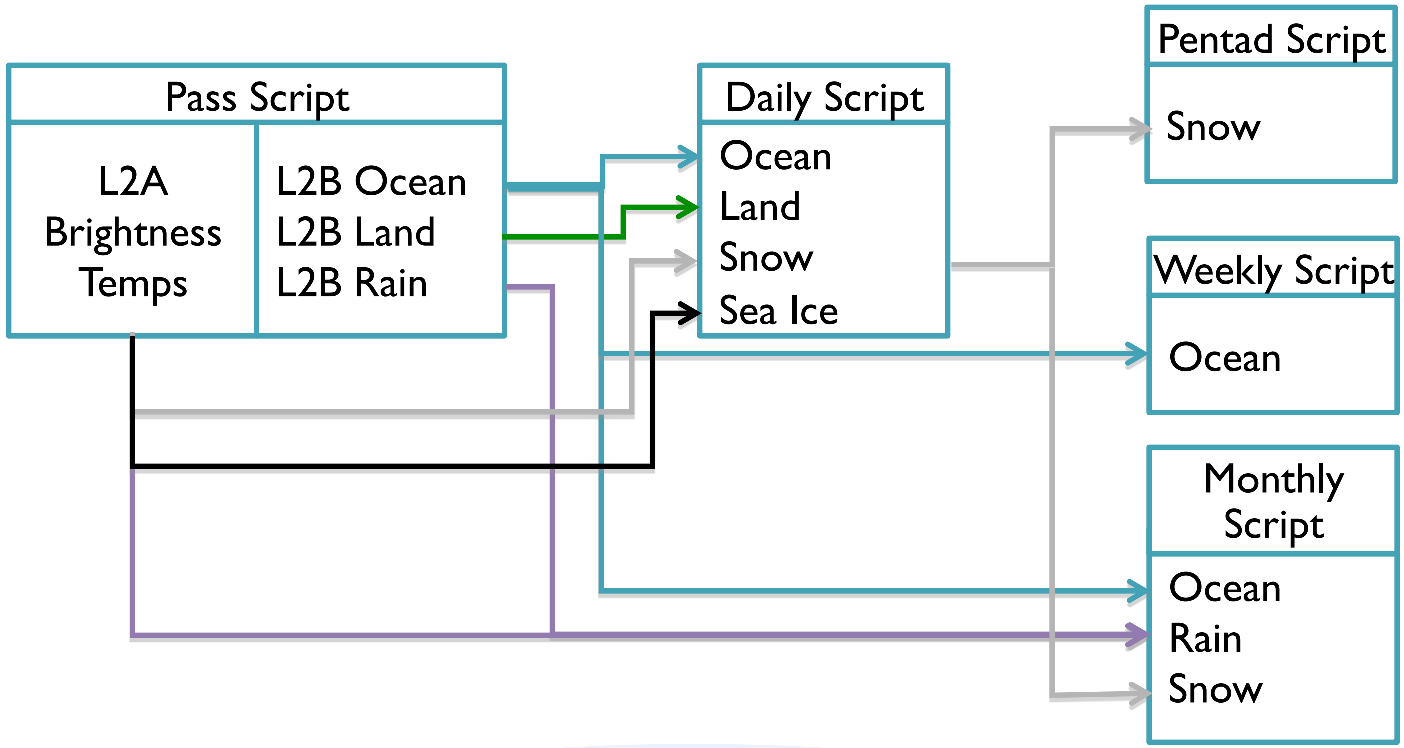
Functional Processing Architecture

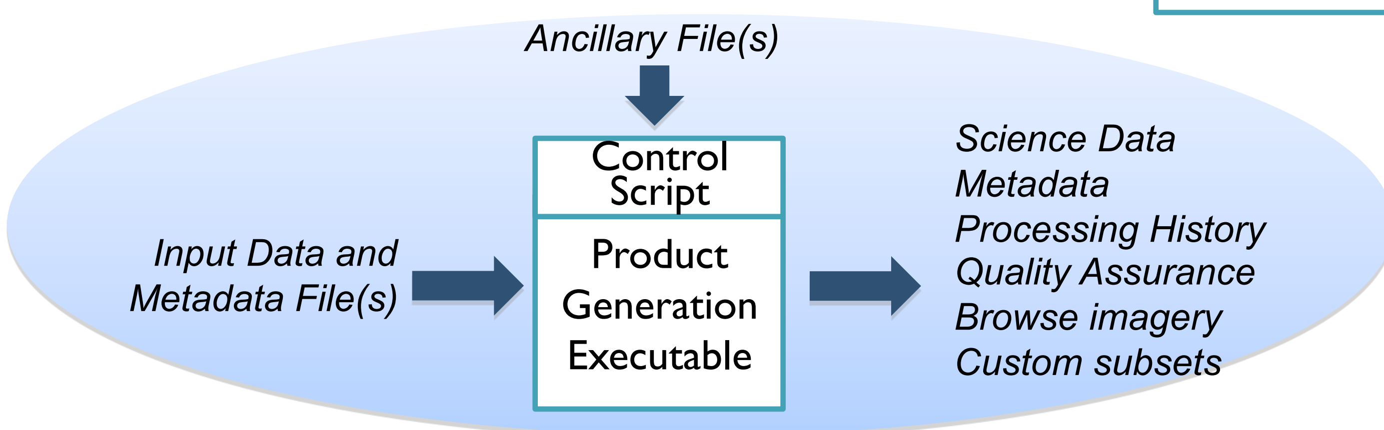


- Ingest: SIPS-GHRC polls SIPS-RSS for new L2A products every 30 minutes
- Product Generation: Routine forward processing is automated to run 24 x 7, unattended
- Storage: All L2A and most recent L2B, L3 products maintained online on one of two 8TB RAID systems
- Delivery: Secure FTP to NSIDC DAAC and AMSR-E Science and Validation teams
- Metrics: Collected from each component and reported to ESDIS

Product Generation

- PGEs provided by the Science Team
- Processing automation controlled by SIPS scripts
  - Pass processing is data driven
  - L3 product generation is scheduled after nominal availability of input products

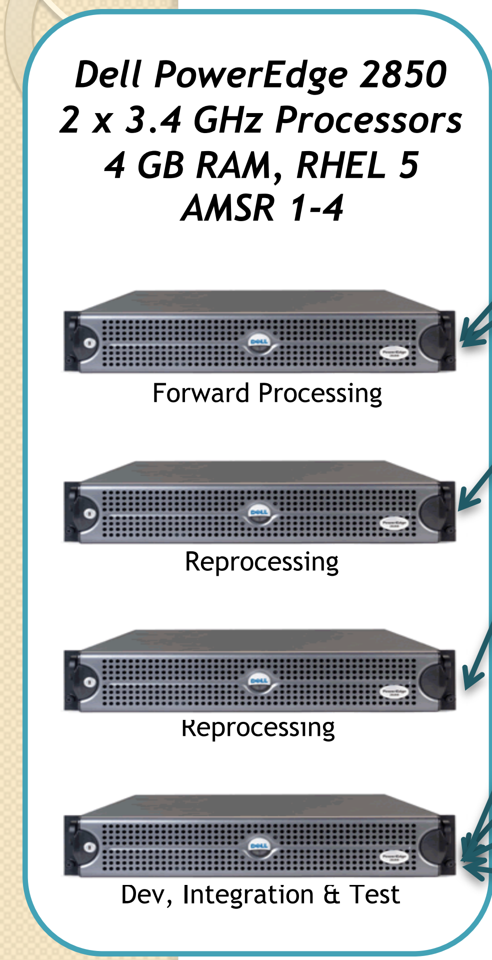




Custom Subsets

- The SIPS generates custom subsets and imagery for the Science Team to support instrument validation and field campaigns
- Database of subset requests includes science product, temporal and spatial bounds
- As products are generated, a final processing step is to create any custom subsets requested for the product

Multiple Processing Environments



- SIPS-GHCC maintains five nearly identical production processing environments
  - Nominal routine forward processing
  - "Catch-up" processing for late arriving Level-2A files
  - Two environments for reprocessing
    - Different time periods for "full reprocessing"
    - Different products for "selective reprocessing"
  - Special processing for science team requests
- Three similarly configured environments support development, integration and testing
  - Development area for SIPS management and processing automation software
  - Science I&T area for testing new PGEs on the SIPS hardware
  - SIPS I&T area for integrating new PGEs with SIPS processing automation software

Flexible Software Architecture

- ***One suite of software can be configured to meet both nominal and off nominal requirements.***
- Forward processing is considered the nominal processing scenario. The development and I&T software configurations follow that model.
- The other three processing scenarios (reprocessing, late, and special processing) have unique requirements that deviate from the nominal. For example:
    - Late arrival of a L2A file triggers reprocessing of L2B and any downstream L3 products which have already been generated
    - Special processing requests normally require only one product type (e.g., Sea Ice or Rain only), and possibly a different version of the PGE

Off Nominal Processing

- Same data processing and management software suite installed in all production environments
- Unique (off nominal) processing requirements are implemented and automated with the help of
  - configuration files and environment variables
  - software switches
  - flexible workflows of atomized operations

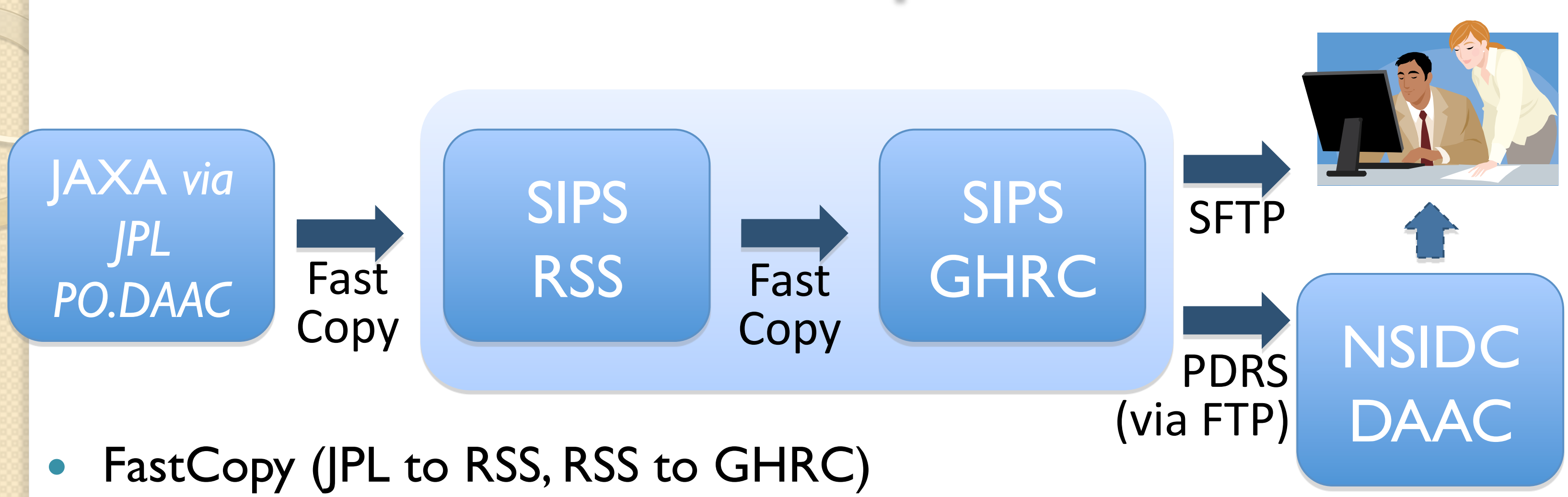
Flexible Processing Workflows

- Encapsulating independent processing operations and combining them into independent low-level scripts enables the SIPS to build custom workflows *for any anomalous situation.*
  - an error recovery scenario could be to "regenerate the Level-3 Land products for April 15 to 30 in their final form, and queue them for delivery to the DAAC"
  - a special processing request might be to "regenerate sea ice subsets for the last three months of 2003 using a different set of latitudes/longitudes"
- Being able to execute the precise software modules needed to deliver the desired products saves time for the SIPS and for the scientists.

Data Storage

- Data products kept on either of two 8TB storage devices
  - Accessible to all processing environments via NSF or GFS
  - Accessible to external users via SFTP
- Data Retention Policy
  - All L2A files from beginning of mission
    - To facilitate reprocessing
    - To serve as offsite archive for NSIDC DAAC
  - L2B and most L3 products for the past 60 days
  - Monthly products for the past 6 months
  - Browse files and custom subsets kept online indefinitely

Different transfer protocols



- FastCopy (JPL to RSS, RSS to GHRC)
  - Commercial software
  - Not actually fast, but very accurate - includes error checking, restart after interruption, etc.
- Secure FTP (SIPS-GHRC to science team)
  - Data access restricted to science team for quality assurance, special products, etc. Other users get data from DAAC.
- Product Delivery Record Server (SIPS-GHRC to NSIDC DAAC)
  - ECS-defined SIPS-DAAC interface
  - Provides for accounting of successful file transfers and any errors

Product Delivery Record Server

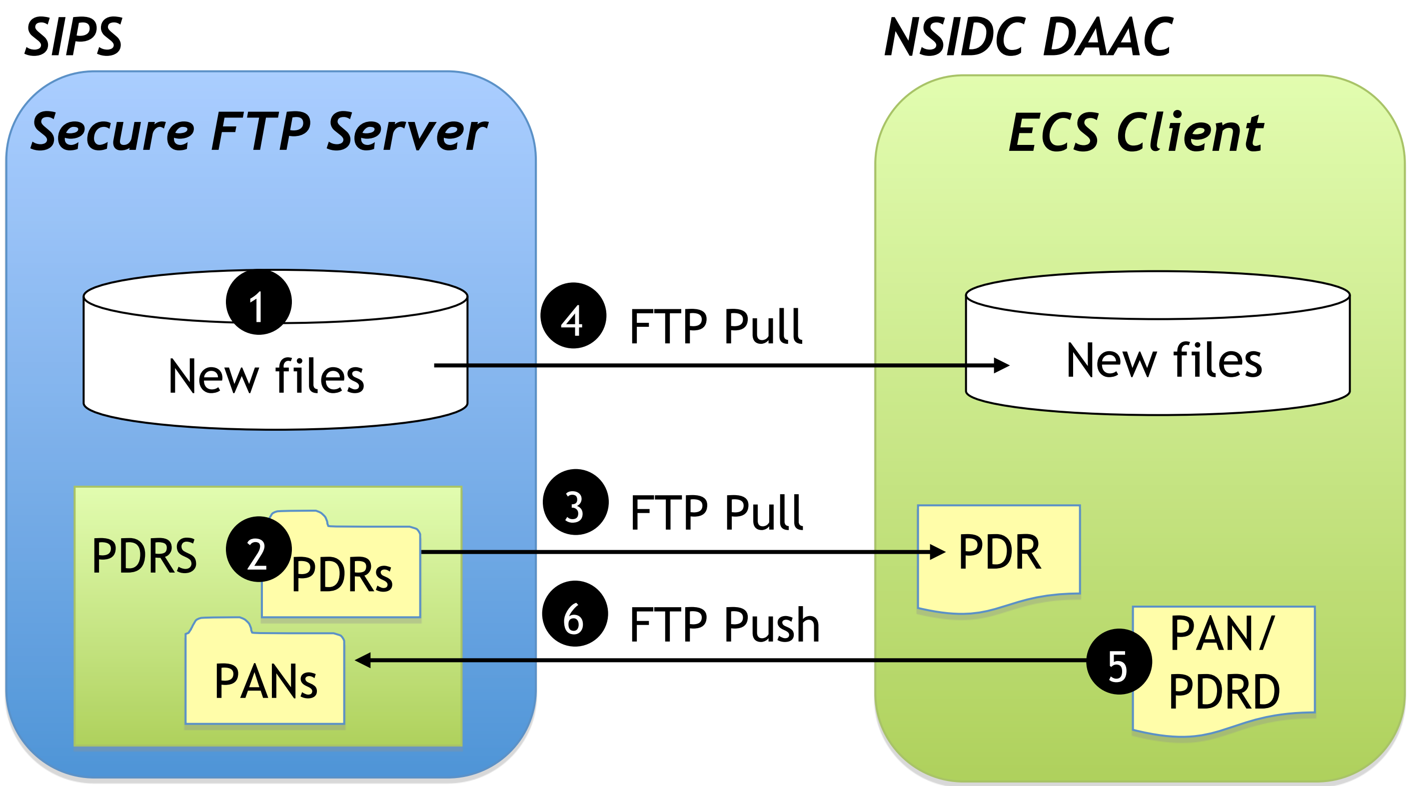
The ICD between SIPS and the EOSDIS Core System specifies

- Product Delivery Record (PDR)
  - Lists files to be delivered by name, location and type,
  - Links all auxiliary files (e.g., metadata, browse) to their parent data file
- Product Acceptance Notice (PAN)
  - Indicates success or failure of data transfer, optionally on per-file basis
  - Depending on the error, failures may be retried automatically or may require human intervention
- Product Delivery Record Discrepancy (PDRD)
  - Indicates syntactic problem with PDR

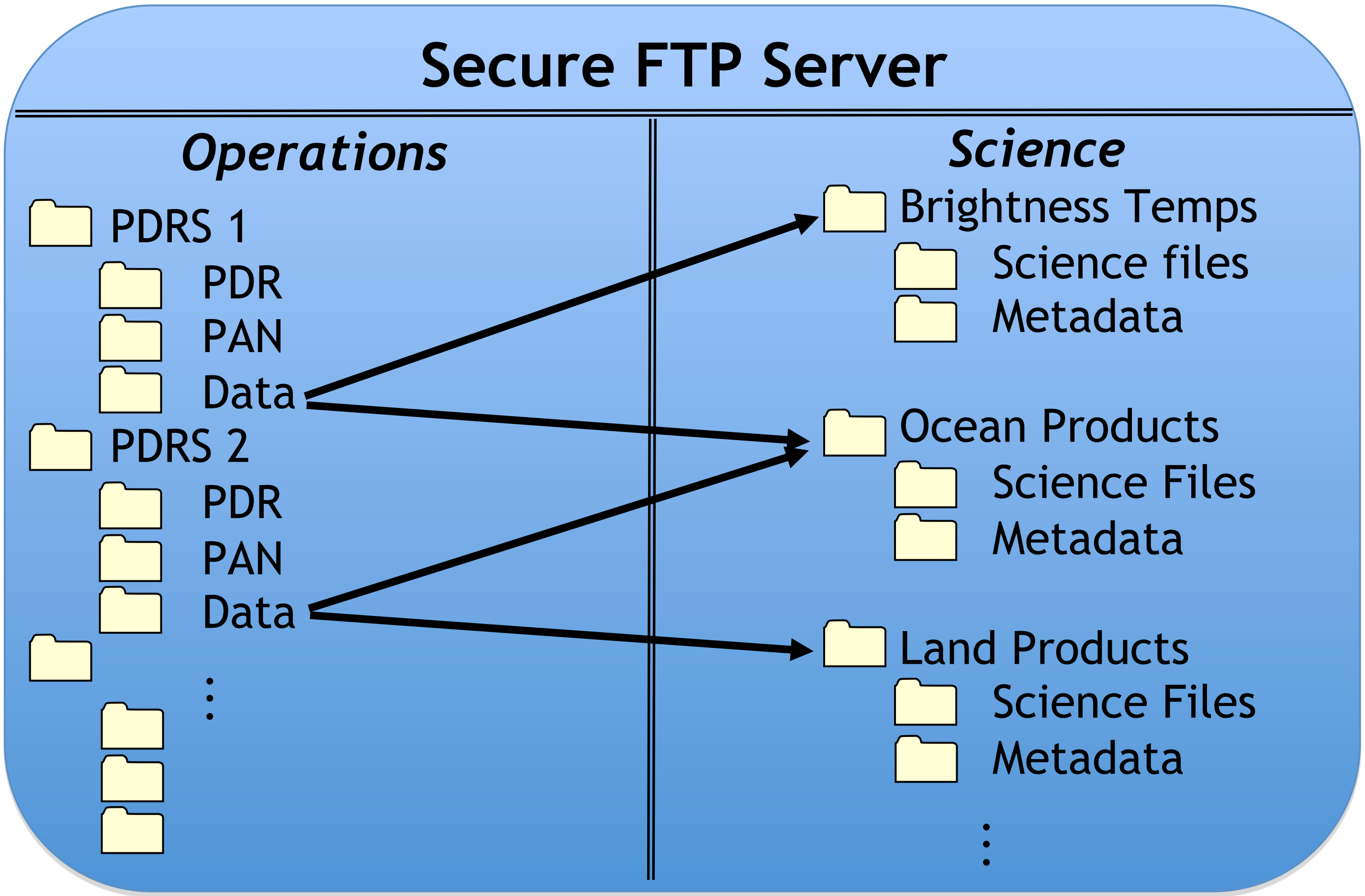
The PDRS software orchestrates all aspects of the data exchange between the SIPS and the DAAC, including

- PDR generation
- file transfers
- processing returned PANs

Basic PDRS Process



FTP Server Architecture



The PAN Handler

- In implementing the PDRS protocol, the SIPS also developed a PAN Handler to:
  - keep a record of all files that were staged for transfer and whether or not they were successfully ingested into the ECS
  - read every PAN to determine what actions, if any, need to be taken based upon the disposition stated in the PAN
  - detect certain classes of errors or failures and automate error recovery
  - move processed PDRs and PANs off the FTP Server, once they are correlated and resolved
  - collect PDRS-related statistics information for metrics reporting

Conclusions

- Standards used
  - Data format: HDF-EOS 2
  - Metadata model: EOSDIS / ECS
  - Data delivery: DAAC-SIPS interface
  - Software: Delivered Algorithm Package
- A single set of automation software that is configurable to meet both nominal and off-nominal requirements enables the AMSR-E SIPS to
  - respond quickly to problems and changing requirements
  - reallocate resources to meet current needs
- The flexible processing framework resulted in reduced cost for long term operations and maintenance, which in turn allows the SIPS to implement additional requirements at minimal cost to the project.